BRAIN RESEARCH 1210 (2008) 204–215

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

# The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study

*Lisa S. Scott*[a,*], *James W. Tanaka*[b], *David L. Sheinberg*[c], *Tim Curran*[d,*]

[a]*Department of Psychology, University of Massachusetts, 413 Tobin Hall/135 Hicks Way, Amherst, MA 01003, USA*
[b]*Department of Psychology, University of Victoria, Canada*
[c]*Department of Neuroscience, Brown University, USA*
[d]*Department of Psychology, University of Colorado, USA*

ARTICLE INFO

*Article history:*
Accepted 13 February 2008
Available online 4 March 2008

*Keywords:*
Perceptual expertise
Categorization
Event-related potential

ABSTRACT

This study examined the neural mechanisms underlying perceptual categorization and expertise. Participants were either exposed to or learned to classify three categories of cars (sedans, SUVs, antiques) at either the basic or subordinate level. Event-Related Potentials (ERPs) as well as accuracy and reaction time were recorded before, immediately after, and 1-week after training. Behavioral results showed that only subordinate-level training led to better discrimination of trained cars, and this ability was retained a week after training. ERPs showed an equivalent increase in the N170 across all three training conditions whereas the N250 was only enhanced in response to subordinate-level training. The behavioral and electrophysiological results distinguish category learning at the subordinate level from category learning occurring at the basic level or from simple exposure. Together with data from previous investigations, the current results suggest that subordinate-level training, but not basic-level or exposure training, leads to expert-like improvements in categorization accuracy. These improvements are mirrored by changes in the N250 rather than the N170 component, and these effects persist at least a week after training, so are conceivably related to long-term learning processes supporting perceptual expertise.


## 1. Introduction

Recent studies of perceptual expertise and categorization have used training studies to further our understanding of the behavioral and neural mechanisms contributing to the acquisition of visual perceptual expertise (Gauthier and Tarr, 1997; Gauthier et al., 1999; Gauthier et al., 1998; Rossion et al., 2002; Rossion et al., 2004; Scott et al., 2006; Tanaka et al., 2005). The use of training studies allows for more precise control over the amount and quality of visual experience needed to obtain perceptual expertise. Although researchers do not expect to be able to equate the acquisition of expertise in the laboratory to real-world expertise, training in a laboratory setting allows for better manipulation of factors contributing to perceptual learning and generalization. Results of perceptual training studies have lead to several important conclusions about how people learn to

* *Corresponding authors.*
E-mail address: lscott@psych.umass.edu (L.S. Scott).
URL: http://people.umass.edu/lscott (L.S. Scott).

0006-8993/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2008.02.054

categorize at different levels, how category learning generalizes to novel exemplars and categories, and how this type of learning may be implemented at the neural level (Gauthier and Tarr, 1997; Gauthier et al., 1998; Rossion et al., 2002; Rossion et al., 2004; Scott et al., 2006; Tanaka et al., 2005).

Tanaka, Curran, and Sheinberg (2005) trained participants to classify species of wading birds and species of owls at either the subordinate (species, e.g., Barn owl) or basic (wading bird) level of abstraction. Training took place on 6 days over a 2-week period with the amount of training trials equated for both subordinate and basic-level conditions. Behavioral results of this study suggest that subordinate, but not basic-level training, increased discrimination of previously trained birds. Moreover, greater generalization to novel exemplars within trained species, and novel exemplars of untrained species (within the same family) was found for subordinate compared to basic-level training. These data suggest subordinate-level discrimination training is an important factor in the acquisition of perceptual expertise and the subsequent transfer to new exemplars from learned categories and new exemplars belonging to novel, but structurally related categories.

In a recent follow-up study, Event-Related Potentials (ERPs) were recorded before and after training at the subordinate and basic levels (Scott et al., 2006). The behavioral results of this investigation replicated previous findings suggesting subordinate but not basic-level training led to increased discrimination of trained birds and increased generalization of untrained birds. We also identified two distinct ERP components, the N170 and the N250, that were correlated with the acquisition of perceptual expertise. Whereas the N170 was sensitive to the encoding of basic-level, shape information, the N250 was modulated by the more fine grain perceptual detail required for subordinate-level identification (Scott et al., 2006; Tanaka et al., 2006). Generalization to untrained exemplars was also found for both the N170 and the N250 components. These results suggest that increased discrimination and generalization required for subordinate-level judgments map more directly onto the N250 component than the N170 component, which has been previously associated with real-world expertise (Tanaka and Curran, 2001; Gauthier et al., 2003). In addition, these data further question the notion that the N170 specifically indexes face processing (Carmel and Bentin, 2002; Eimer, 2000; Sagiv and Bentin, 2001) and instead provide additional evidence for a more general experience based N170.

The present investigation sought to further clarify the factors contributing to the acquisition of perceptual expertise, including the function of the ERP components correlated with categorization and perceptual expertise. This research addresses three unanswered questions. First, how does learning, mediated by tasks including feedback and category labeling, differ from exposure-only learning? Previously, we found both behavioral and electrophysiological differences for categories of birds trained at the subordinate versus the basic level (Scott et al., 2006). Here we extend this finding and examine whether subordinate- and basic-level learning contribute anything above and beyond simple exposure learning. Second, for how long after training are behavioral and electrophysiological training effects maintained? More specifically, is it necessary for training to continue in order to sustain the increases in performance and ERP amplitude we previously reported (Scott et al., 2006)? If the effects of training are short-lived in this paradigm, we must consider the relevance of these results to real-world perceptual expertise. Finally, does training with cars, an artificial (as

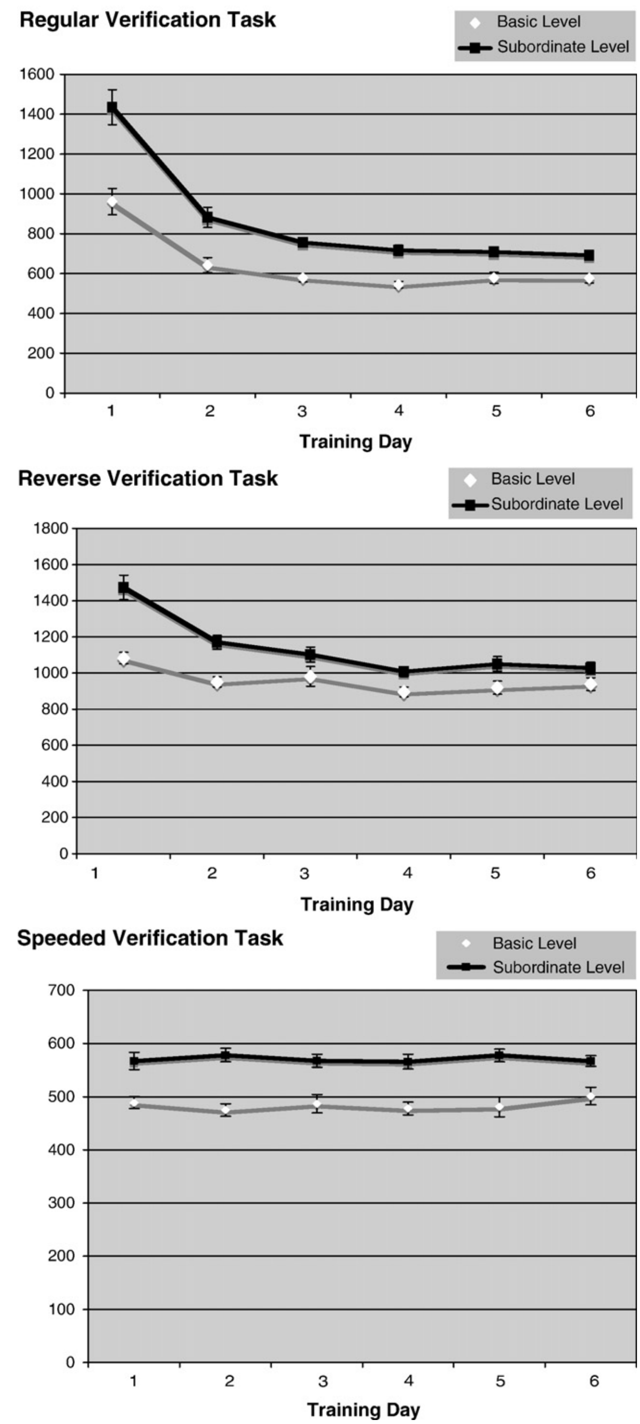



**Fig. 1 – Reaction time (mean and standard errors) changes across training. Behavioral data illustrating an entry-level shift in reaction time (RT) across both the Regular- and the Reverse-Verification Tasks. No entry-level shift was observed for the Speeded-Verification Task.**

opposed to a natural kind) object category, influence learning and generalization across different levels of training? Multiple exemplars of multiple models of three different types of cars were used as experimental stimuli. Car stimuli were used to first determine whether or not we could replicate and extend our previous results with a new class of stimuli. Furthermore, we wanted to establish whether learning objects from a human-made category, such as cars, yielded similar or different results than from learning objects from a natural category, such as birds.

## 2. Results

### 2.1. *Behavioral results*

Due to large variability between the numbers of completed blocks across subjects (see Experimental procedures), analyses were not conducted for the naming task. During the subsequent training tasks, reaction time (RT) measures were used to monitor the effects of training. RTs were computed for correct responses only (see Fig. 1). Accuracy across all training days for all tasks was at or near ceiling.

The category verification tasks were analyzed to determine whether RT's changed across 6 days of training. Reaction time data from 2 subjects was excluded due to experimenter error and loss of data. These analyses revealed greater overall RT's on trials requiring a subordinate-level verification compared to a basic-level verification for all three tasks: regular verification ($F(1,9)=85.54$, $p=.0001$); reverse verification task ($F(1,9)=46.19$, $p=.0001$); and speeded verification ($F(1,9)=79.02$, $p=.0001$). Greater RT's were also found on the first day of training compared to all other days for the regular verification ($F(5,5)=7.56$, $p=.022$) and reverse verification ($F(5,5)=23.39$, $p=.002$) but not for the speeded verification task. A significant interaction between training type and day was found for the regular verification ($F(5,5)=48.65$, $p=.0001$) and reverse verification ($F(5,5)=7.41$, $p=.023$) tasks, but not for the speeded verification task. An examination of the means suggest that subordinate-level performance became increasingly similar to basic-level performance across days. However, on the last day of training, RT to basic-level trials was

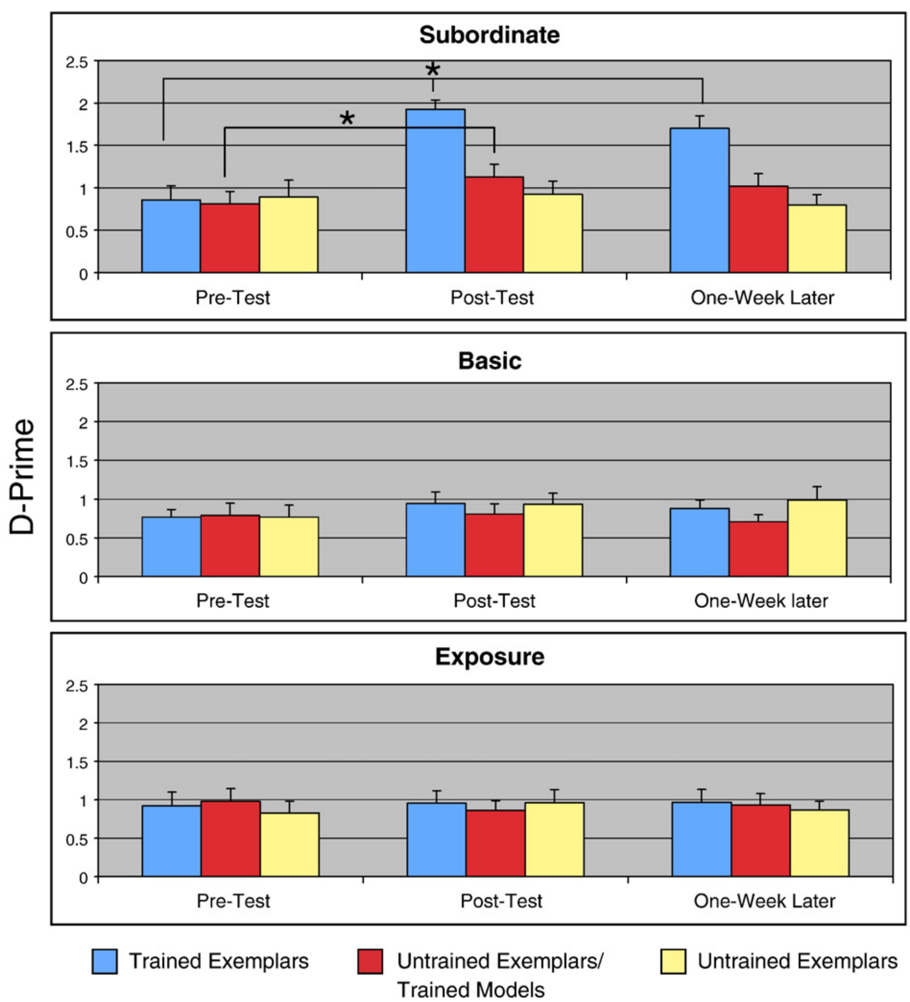



**Fig. 2 – Pre-, post-, and 1-week post-training matching performance. Behavioral d' scores (mean and standard errors) across all conditions for the sequential matching task. Brackets indicate significant differences within categorization training type.**

still significantly faster than RT to subordinate-level trials across all tasks ($p$'s < .001). Fig. 1 depicts this effect for the regular, reverse, and speeded verification tasks.

Subordinate-level discrimination performance was assessed before, immediately after, and 1-week after training across trained and untrained exemplars and models using a successive matching task. Subordinate-level discrimination of cars trained at the subordinate and basic level, as well as exposure-only trained cars were tested in this task (Fig. 2). Accuracy increased from pre-test (72.2%; SD = 2.6%) to post-test (78.1%; SE = 2.0%) and remained constant from the immediate post-test to the 1 week post-test (77.2%; SE = 1.7%). Planned comparisons, investigating the behavioral effects of training, reveal a significant increase in d' from pre-test to post-test for exemplars of cars trained at the subordinate level ($t(11) = -6.5$, $p = .0001$). Subordinate-level training generalized to untrained exemplars of trained car models ($t(11) = -2.7$, $p = .019$) but not to exemplars of untrained models. There was no evidence of a d' increase from pre- to post-test for cars trained at either the basic level or for the exposure-only condition. Comparisons of pre-test measures of d' to 1-week post-test measures of d' reveal a increase for exemplars of trained cars ($t(11) = -5.2$, $p = .0001$), but no evidence of generalization at the 1-week post-test. Although still significantly greater than pre-test levels, a d' decrease was found from post-test to 1-week for exemplars of cars trained at the subordinate level ($t(11) = 2.67$, $p = 022$). In addition, d' for the untrained exemplars of trained car models was greater with subordinate-level training compared to basic-level training for the 1-week post-test ($t(11) = 2.45$, $p = .032$).

## 2.2. Electrophysiological results

### 2.2.1. N170

Subordinate-level, basic-level, and exposure-only training increased N170 amplitude in a manner that generalized across all conditions ($F(2,10) = 16.03$, $p = .001$; see Figs. 3 and 5). Follow-up analyses of this effect reveal a greater N170 response at the initial post-test compared to both the pre-test ($p = .001$) and compared to the 1-week post-test ($p = .028$). There was no mean amplitude difference between the N170 response in the pre-test versus the 1-week post-test. There were no latency differences.

### 2.2.2. N250

Analyses reveal a main effect of test ($F(2,10) = 10.43$, $p = .004$) and a main effect of stimulus presentation order ($F(1, 11) =$

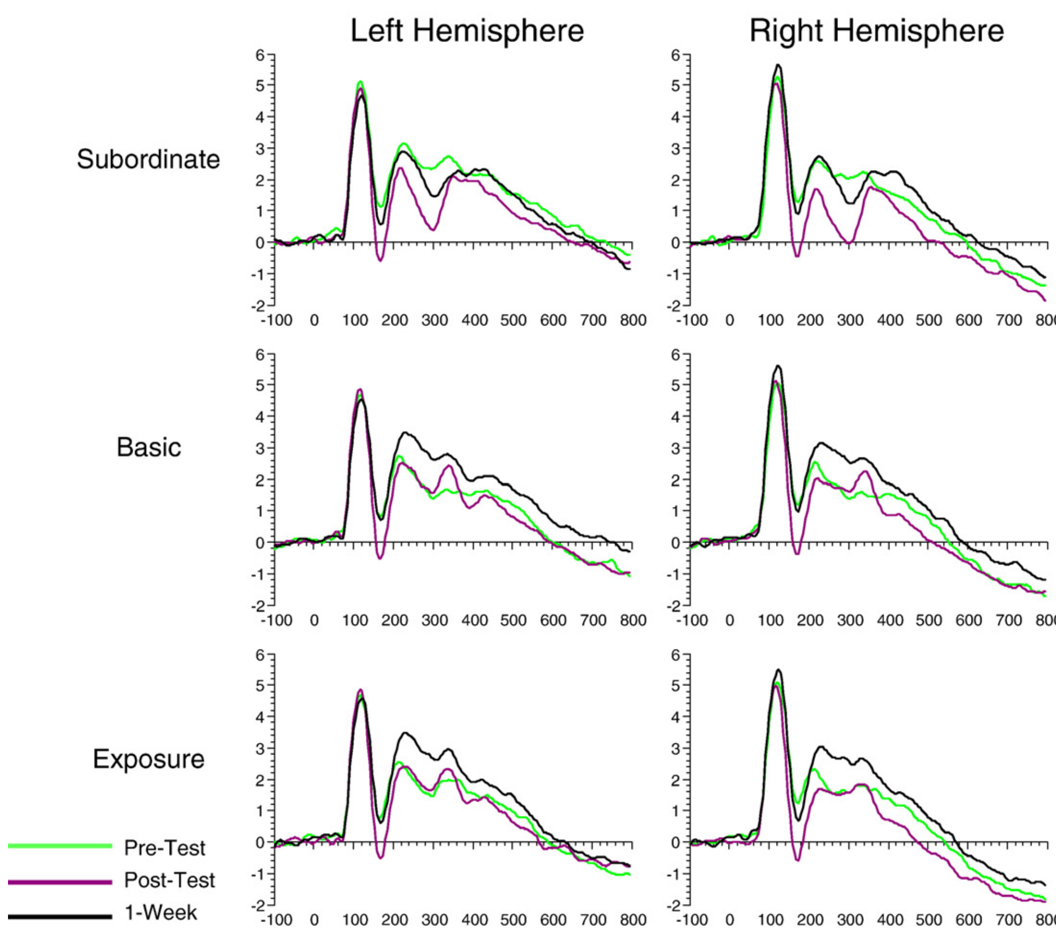



**Fig. 3 – ERP waveforms. The graphs in the left column represents an average of electrodes in the left hemisphere (64, 65, 66, 69, 70, 71, 74, and 75) and the right column represents an average of electrodes in the right hemisphere (83, 84, 85, 89, 90, 91, 95, and 96) across the three training types (Subordinate, Basic, Exposure). The increase in the N170 across all three training types at Post-test is best seen in this figure.**

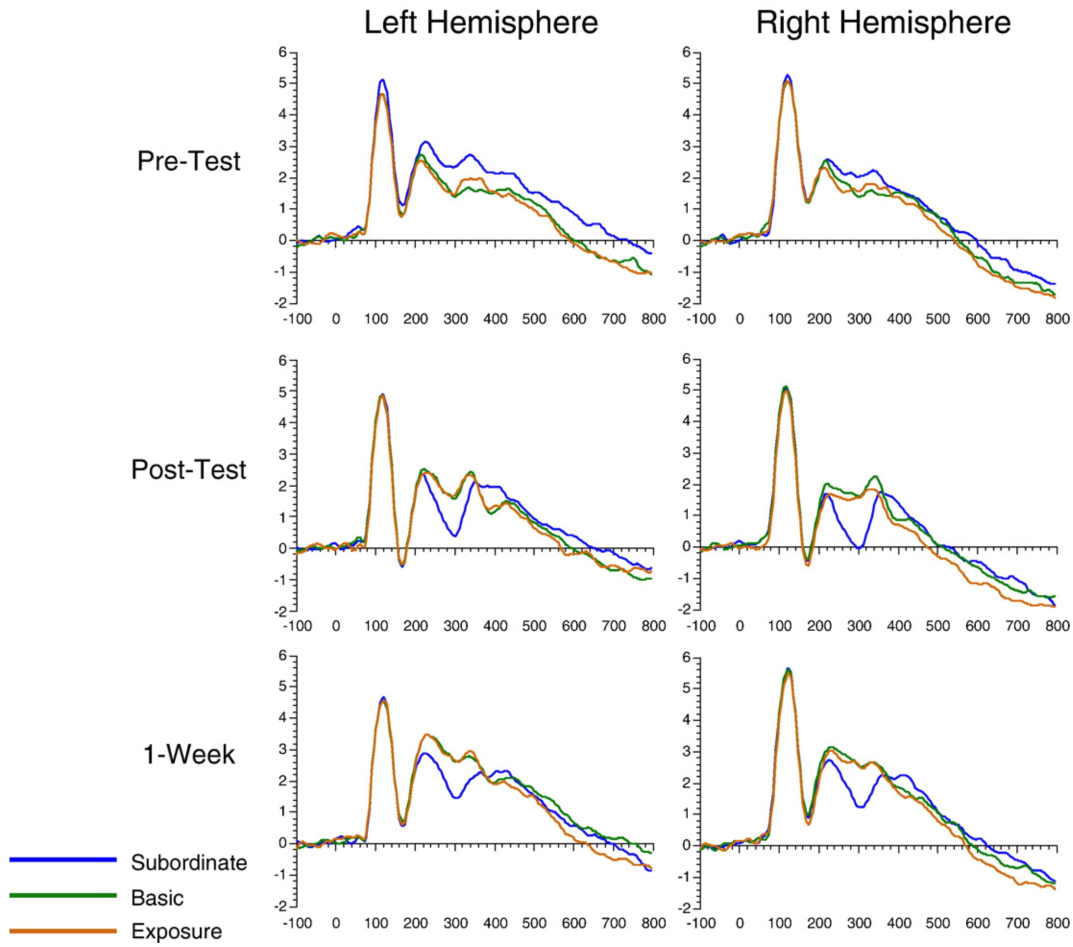



**Fig. 4 – ERP waveforms. The graphs in the left column represents an average of electrodes in the left hemisphere (64, 65, 66, 69, 70, 71, 74, and 75) and the right column represents an average of electrodes in the right hemisphere (83, 84, 85, 89, 90, 91, 95, and 96) across the three testing sessions (Pre-test, Post-test, 1-week). The increase in the N250 for subordinate trained cars at both the post-test and the 1-week post-test is best seen in this figure.**

10.70, $p=.007$). Follow-up pairwise comparisons and an examination of the means suggest an overall greater N250 response at the post-test compared to the 1-week post-test ($p<.01$) and a greater N250 to the first presentation of a stimulus compared to the second presentation within a trial. These main effects are qualified by several interactions.

First, an interaction between test day and categorization training type was found ($F(4,8)=4.15$, $p=.041$, see Figs. 4 and 5). Prior to training, the N250 did not significantly differ between the three training conditions, ($p$'s$>.05$). However, at the post-test as well as 1 week later, the N250 was more negative for subordinate than basic or exposure training. These subordinate training effects generalized across all conditions and were present at both the immediate post-test ($p$'s$<.05$) and the 1-week post-test ($p$'s$<.05$).

There was also an interaction between stimulus presentation order and hemisphere ($F(1, 11)=5.50$, $p=.039$). An examination of this interaction revealed that there is a greater N250 for the first presentation, compared to the second presentation, in the right, but not the left, hemisphere ($p<.05$). There were no latency differences.

### 2.2.3. *Dipole Source Analysis*

Source estimation was performed for the N170 and N250 using BESA (Version 5.1.8). Based on previous source estimation results (Scott et al., 2006) source analyses were conducted for the difference of the pre-test subordinate condition and the post-test subordinate condition for both the N170 and the N250 using a component onset-to-peak window (N170=148–184 ms; N250=232–280 ms). Spatial principal components analysis (PCA) revealed one factor for the N170 (99.1% of the variance explained) and one factor for the N250 (99.0% of the variance explained); therefore one pair of laterally symmetric sources was fitted for each component. For the N170, the Talairach coordinates for the center of activity was $x=\pm 31$, $y=-47$, $z=15$ (residual variance (RV)=6.5%, see Fig. 6 left). For the N250, the Talairach coordinates for the center of activity was $x=\pm 15$, $y=-45$, $z=21$ (RV=7.7%, see Fig. 6 right). Given the similarity of these two locations, the N170 solution also provided a reasonably close fit to the N250 (RV=13.8%). Both of these locations correspond to white matter tracks within the posterior/superior temporal lobe, which are unlikely to be the true sources of the postsynaptic potentials generating

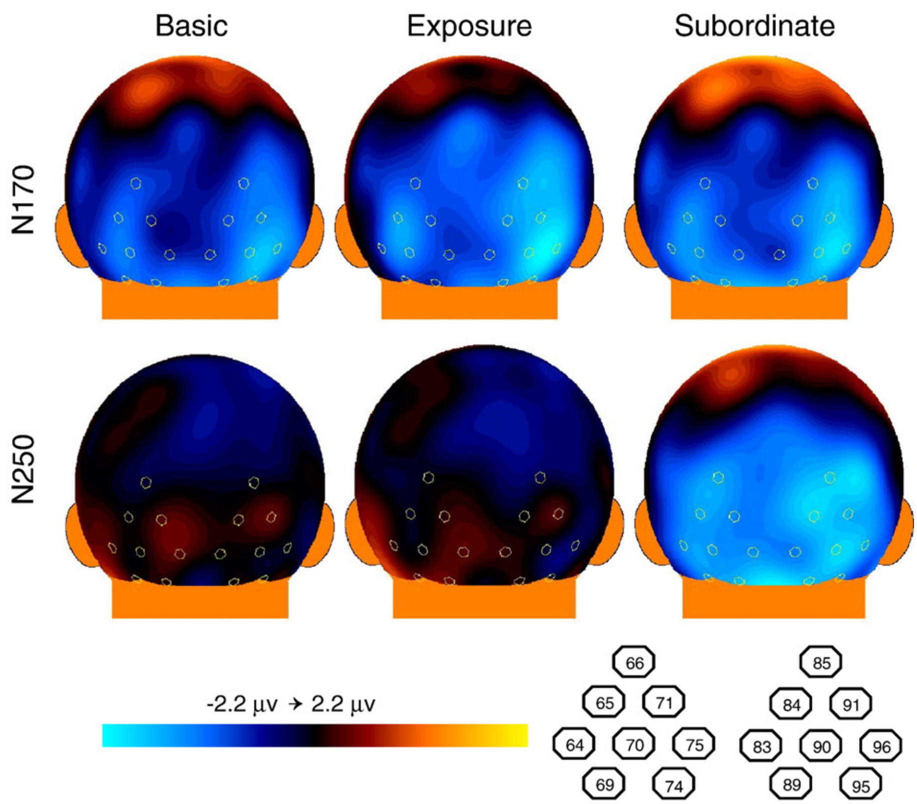



**Fig. 5 – Topography of Subordinate-level effects. Topographic map of the difference between the pre and post subordinate-level for the N170 (155–211 ms) and the N250 (230–330 ms) across the three training conditions. The electrode location numbers are highlighted in the bottom right-hand corner.**

these ERPs, but the close proximity of these estimates does not provide strong evidence for the anatomical separability of the N170 and N250 sources.

## 3. Discussion

The behavioral results of this investigation replicate previous findings showing that subordinate-level training leads to increased discriminability among trained car exemplars. This training effect persisted 1 week after the end of training. Similar to our previous work using bird stimuli (Scott et al., 2006; Tanaka et al., 2005), we found that learning generalized to untrained exemplars of trained car models. However, unlike the studies with birds, car training did not lead to generalization of untrained car models. The electrophysiological results suggest that mere exposure, basic-level training, and subordinate-level training all lead to significant increases in N170 amplitude, but that this effect is not maintained 1 week after training. Furthermore, these data replicate the results of Scott et al. (2006), in that the N250 was found to index subordinate-level access to objects. Unlike the N170, the increased N250 amplitude is maintained 1 week after training ends. Together with data from previous investigations, the current results lead us to conclude that subordinate-level training, but not basic-level or exposure training, leads to expert-like improvements in categorization accuracy; and that these improvements are mirrored by changes in the N250 rather than the N170.

Previous research has found an enhanced N170 when participants view faces (e.g. Carmel and Bentin, 2002; Eimer, 2000) or objects of expertise (e.g. Gauthier et al., 2003; Tanaka and Curran, 2001). The findings of the present study suggest that this enhancement of neural processing occurs because of the increased level of consistent exposure people have to faces as well as other objects of expertise. Our results show an increased N170 amplitude response regardless of whether the participants were trained at the basic, subordinate, or exposure-only levels. However, the enhanced N170 that was present immediately after training was short-lived and was no longer evident 1 week later in any of the three training conditions. Combined with our previous investigation using bird stimuli (Scott et al., 2006) this finding suggests that larger N170 responses are due to an increase in category exposure, which must be maintained over time. This neural increase also generalizes to previously untrained exemplars of trained cars as well as previously untrained models of cars. Thus, previous findings of increased N170 amplitude associated with faces (e.g. Carmel and Bentin, 2002; Eimer, 2000) and other objects of expertise (e.g. Gauthier et al., 2003; Tanaka and Curran, 2001) are likely to reflect greater categorical exposure to these stimuli rather than expert identification per se. We previously reported a link between the N170 and basic-level categorization and the N250 and subordinate-

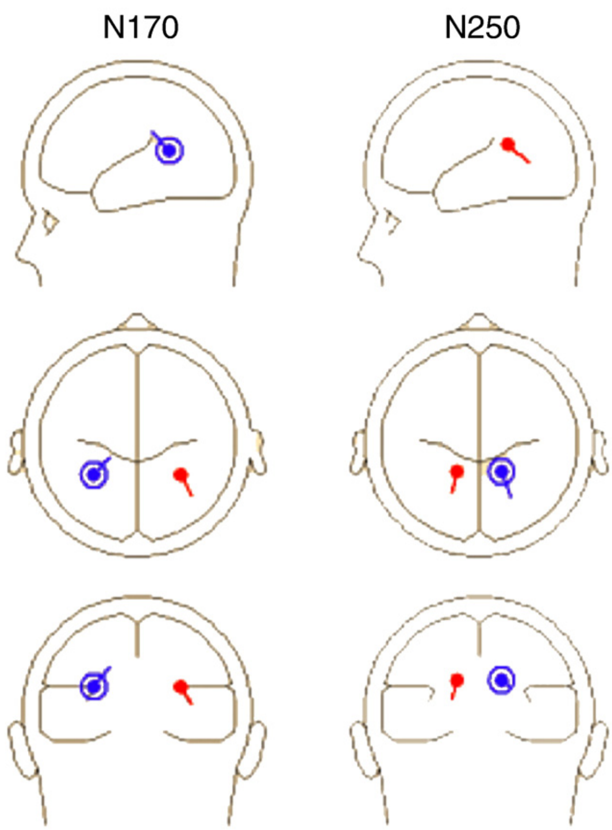



**Fig. 6 – Source modeling. Panels are estimates of source localization for the difference of the post-test subordinate-level minus the pre-test subordinate-level condition for each component.**

level categorization (Scott et al., 2006). However, given the lack of differences between basic-level and exposure-only training reported here, there now appears to be no reason to link the N170 to explicit basic-level categorization. It is possible that the N170 reflects a lower-level categorical matching process, which may be modulated by previous categorical exposure or unsupervised category learning. On the other hand, it is also possible that participants spontaneously categorized exposure stimuli at the basic level, even though they were not instructed to do so. More work is needed to test these possibilities.

The present N170 results are also consistent with another recent training study investigating rapid adaptation fMRI changes before and after car categorization training (Jiang et al., 2007) and to neural recordings from non-human primates (Freedman et al., 2006; Anderson et al., in press). Jiang et al. (2007) report that training led to a sharpening of the stimulus representation in the lateral occipital complex (LOC) which appeared to not be related to any specific category training or task, but instead to experience with the physical category shape. The authors conclude that this supports a model of category learning involving two mechanisms, the first, a shape based but task-irrelevant representation, that then feeds into the neural circuits involved in later categorization. Similarly, Freedman et al. (2006) report that passive exposure to categories of stimuli as well as explicit training with category exemplars both lead to increased selectivity of single cells in inferior temporal cortex of monkeys. Anderson et al. (in press) extended these findings by showing significant experience dependent increases in local field potential amplitude in monkey temporal cortex, with differences between novel and familiar stimuli first detectable approximately 170 ms after stimulus onset. The results reported here fit nicely with these findings and suggest that the N170 might reflect the electrophysiological analog of this experience based, shape-specific, representation.

ERP research has uncovered a variety of negative components peaking between 200 and 350 ms after stimulus onset, but as previously indicated, the N250 seems to be the best designation for the second of our primary ERP effects. Anterior N2 components have previously been found to be involved in cognitive control, the detection of novelty, and orienting (see Folstein and Van Petten, 2008 for a review of N2 components). The N2b, typically largest over central electrode locations for auditory stimuli and over posterior sites for visual stimuli is observed in odd-ball type tasks when the deviant stimuli are task-relevant (Simson et al., 1977). Another posterior N2, the N2pc has been found in visual search paradigms and is elicited by targets presented in the contralateral visual field (Luck and Hillyard, 1994). A third posterior N2, the N2pb is a bilateral response that is sensitive to stimulus probability (Luck and Hillyard, 1994). A separate posterior N2, the N250 is involved in aspects of visual processing, and has previously been found to be larger in response to the repetition of familiar relative to unfamiliar faces (Schweinberger et al., 2002; Schweinberger et al., 2004; Schweinberger et al., 2002; Tanaka et al., 2006) and is larger in response to bird stimuli trained at the subordinate-, relative to the basic-level (Scott et al., 2006). In the present investigation, we observed a posterior N2 in a sequential matching task, which does not vary stimulus probability or require target detection (all items are targets). We argue that the N2 that we observe is the N250, as studied by Schweinberger et al. (2002) rather than the other negativities related to target detection and stimulus probability. However it is possible that participants are treating the subordinate-level trained stimuli as targets implicitly, thus we cannot rule out the possibility that this posterior component contains subcomponents related to target detection (see Appendix A for a sampling of waveforms across the scalp).

In the current experiment, unlike the N170, the N250 response (shown in Figs. 4 and 5) increases only to cars trained at the subordinate level. Neither basic-level nor exposure-only training influences the N250 response. This finding replicates our previous report (Scott et al., 2006) and is consistent with studies investigating the N250 in response to face stimuli (e.g. Schweinberger et al., 2004; Tanaka et al., 2006). The face N250 component was previously demonstrated to be larger in response to familiar relative to unfamiliar faces (Tanaka et al., 2006; Schweinberger et al., 2002, 2004). We argue here that this effect is not face-specific and is likely due to the increased subordinate-level access to familiar faces.

Similar to the N170, the N250 response also generalizes to the untrained stimuli, however unlike the N170, the N250 increases are seen at post-test and at 1-week post-test. The longevity of this effect suggests that the N250 may be a potential marker of the long-term learning processes that underlie perceptual expertise rather than a temporary effect of training.

Combined, the electrophysiological results suggest that the N170 component may not be specifically related to perceptual expertise and may instead be a byproduct of the fact that experts typically see objects of expertise more often than novices. The N250 seems to more clearly index learning occurring at the subordinate level, which is typical of perceptual expertise. Because only subordinate-level training influenced matching task accuracy, the processes underlying the N250 seem more likely related to these behavioral improvements than those underlying the N170.

One important difference between the current investigation and our previous training study (Scott et al., 2006) is lack of behavioral generalization to novel exemplars from untrained models. These effects are somewhat surprising given the previous findings showing transfer effects for both untrained exemplars of trained species of birds and untrained species of birds (Scott et al., 2006; Tanaka et al., 2005). However, there are important differences between birds and cars that might account for differences in transfer. As natural kinds, common species of birds (e.g., screech owls, barred owls, spotted owls) bear a physical resemblance to one another owing to their shared genetic makeup. In contrast, because cars are human-made objects, exemplars from the category of SUV, antique or sedan cars are less constrained in their appearance and might be less structurally similar. That is, unlike species of birds belonging to a common avian family (e.g., Great Grey Owl, Screech Owl), models of cars belonging to the same class of car (e.g., Chevy Tahoe, Honda Pilot) are less likely to structurally similar. Because high within-category similarity promotes generalization to novel exemplars more than low within-category similarity (Posner and Keele, 1968; Homa and Vosburgh, 1976), category transfer might be more likely for novel exemplars from the more homogenous bird category than for novel exemplars from the more heterogeneous car category.

It was informative that transfer effects were found in the untrained exemplar/trained model condition but not in the untrained exemplar/untrained model condition. The untrained exemplar/untrained model condition constitutes a more demanding test of transfer because participants must generate a new category representation (i.e., new car model) in response to the unfamiliar input stimulus. By contrast, the untrained exemplar/trained model condition only requires that novel stimulus is associated with a pre-existing category representation. Whereas *intermediate* perceptual transfer only requires the activation of a familiar category representation from a novel input, *strong* transfer demands the construction of a new category representation and therefore, provides a more stringent test of transfer (Williams and Tanaka, in press).

Scott et al. (2006) used dipole analysis to argue that the N170 and N250 most likely originated from different anatomical sources. In the present experiment we report very similar N170 and N250 sources (see Fig. 6), consistent with the topographic similarity of the effects seen in the rightmost column of Fig. 5. Based on source differences, we previously argued for the possibility of qualitatively different processes underlying basic- versus subordinate-level categorization based on spatiotemporal ERP differences (Scott et al., 2006). However, we also allowed for the possibility of a single-process mechanism, differing only quantitatively, but at different times during processing. The results of the present investigation are more consistent with the view that different levels of categorization are the results of a single mechanism, separated by quantitative and temporal differences. This interpretation is consistent with a recent failure to find qualitative differences between basic and subordinate-level processing using a speed-accuracy trade-off task (Mack et al., 2007) and Riesenhuber and Poggio (2000, 2002) suggestion that subordinate-level discrimination requires a more fine-grained perceptual analysis than basic-level discrimination (Collin and McMullen, 2005).

Overall, the present study makes four important contributions to our understanding of category learning and perceptual expertise. First, it replicates previous research showing that subordinate-level training leads to better discrimination of exemplars within categories. Second, these results suggest that subordinate-level training compared to basic-level and exposure-only training differentially influenced the N170 and the N250 ERP components. Specifically, the N250 appears to be influenced by training at the subordinate level, whereas all three types of training equally affected the N170. Third, this investigation is the first to examine retention effects in expertise both behaviorally and electrophysiologically. Our results suggest that subordinate-level training, rather than basic-level or exposure training, leads to increased performance and increased N250 amplitude and that these changes persisted at least 1 week after training. However, the N170 increase seen immediately after training is not maintained 1-week post-training. These findings suggest that subordinate-level learning and the N250 are related to processes involved in the acquisition of long-term perceptual expertise. Finally, for cars trained at the subordinate-level, there appear to be no source or topographic differences between the N170 and the N250, despite temporal differences. This finding supports quantitative differences between the processes underlying these two components.

## 4. Experimental procedures

### 4.1. Participants

Participants included 19 right-handed, undergraduates recruited from the University of Colorado at Boulder. All participants gave informed consent to participate in this study. One subject was excluded due to failure to complete all sessions. Six subjects were excluded due to programming error that resulted in the omission of one of the experimental conditions. The final sample included 12 participants (6 female).

Each participant completed 8 sessions on different days within a 2-week period and 1 additional session 1-week after the 8th session, for a total of 9 sessions. ERPs were recorded on the first, 8th, and 1-week after sessions. Subjects were paid $15/h for ERP sessions, $10/h for behavioral training sessions, and were paid a bonus of $20 for completing all 9 sessions.

### 4.2. Stimuli and apparatus

Stimuli were full color digitized photographs of 3 classes of cars: modern SUVs, modern sedans, and antique cars obtained from various websites. Stimuli included twelve exemplars of twenty different models of SUV's, twenty different models of sedans, and twenty different models of antique cars. The training set was composed of six exemplars of ten different models of each of the classes of cars. The test set of stimuli included the trained exemplars, untrained exemplars of trained models, and exemplars of untrained models. For each of these test conditions, 60 exemplars were obtained by selecting 6 exemplars in each of 10 models. Lures in the training tasks included 30 exemplars of classic cars (models chosen randomly). Stimuli were counterbalanced such that four participants were trained at the subordinate with Antique cars, the basic level with Sedans, and were exposed to SUVs; four participants were trained at the subordinate level with SUVs, the basic level with antique cars, and were exposed to Sedans; and four participants were trained at the subordinate level with Sedans, the basic level with SUVs, and were exposed to antique cars. Within each of the training and test sets, stimuli were pseudo-randomly counterbalanced. For example, all four participants trained at the subordinate level with SUVs were trained with a different set of exemplars. The images were cropped to show only the car and were placed on a white background. All stimuli were 163–292 pixels wide and 89–220 pixels high and were presented at a visual angle of 4.01–4.58° horizontal by .097–3.55° vertical. Stimuli were displayed on a 15-in Mitsubishi flat-panel monitor.

### 4.3. Procedure

All procedures were approved by the Institutional Review Board at the University of Colorado and were conducted in accordance with this approval.

### 4.4. Electrophysiological pre and post-training assessment

Before, immediately after, and 1 week after training, participants completed a subordinate-level sequential matching task that has previously been shown to be sensitive to differences in levels of perceptual expertise (Gauthier et al., 2000) and successfully used during EEG recording (Scott et al., 2006). Participants were shown a stimulus for 800 ms followed by a fixation point for 800–1000 ms and then another image for an additional 800 ms. Then the participants were immediately presented with a question mark and were required to indicate whether the two images were of the SAME (e.g. two Honda CRV's) or of a DIFFERENT (e.g., a Honda CRV & a Toyota Rav 4) model. The question mark remained on the screen until a response was made. SAME trials were always different exemplars of the same model of car. DIFFERENT trials included two exemplars of different models within the same class. Different trial lures were selected randomly (without replacement) from a pool of all other models within the same class of cars. This task consisted of a total of 540 trials. One hundred and eighty trials were SUV's, 180 were sedans, and 180 were antique cars. To monitor changes related to training, the same stimuli were included in the pre-test, post-test, and 1-week-later tasks. All stimuli were randomly ordered and randomly matched within each condition. Across both same and different trials, there were three different types of trials: 1) Trained Exemplars/Trained Models, 2) Untrained Exemplars/Trained Models and 3) Exemplars of Untrained Models. The trained exemplars/trained models condition included 60 images from the training sessions, the untrained exemplars/trained exemplars included 60 new pictures of the trained models, and finally the exemplars of untrained models included 60 images of new models of cars that were never trained. Assignment of exemplars to test conditions was counterbalanced across subjects.

### 4.5. Behavioral training tasks

After the pre-training assessment, participants completed 6 days of training. Stimulus classes (SUV's, sedans, antiques) were counterbalanced across training conditions (subordinate-level, basic-level, and exposure-only). For example, one subject was trained at the subordinate level with SUV's, the basic-level with sedans, and exposure training with antique cars. Within each subject the number of training exposures was equated across the three training conditions. Within each session the first training task was a naming task, and the second a category verification task.

1) *Training Task 1 (Naming)*: Participants learned to label different models of SUV's, sedans, or antique cars. Participants completed 9 blocks of naming training on each day. During the subordinate and basic-level training, participants were first shown 2 models of each class of car (i.e. 2 SUV's and 2 sedans) and increased by 1 more model/class every time they got a block of trials correct. All ten models were trained on each day. The exemplars rotated across blocks and days, and all trained exemplars were presented on each day. The first presentation of each model was labeled, for example a Toyota 4 Runner with either the subordinate-level label "This is Model Y" or the basic-level label "This is Other." Arbitrary labels, rather than actual model/class names where used to help reduce the effects of prior knowledge. For the subordinate-level training, participants then pressed the "Y" key whenever they saw a Toyota 4 Runner. For the basic-level training participants pressed the "O" key whenever they saw a sedan. The label was only present for the first presentation of each model in each block. Participants were required to score 100% in each block to move on; otherwise, the block was restarted. Feedback, including the correct answer, was given for 1500 ms for incorrect responses.
2) *Training Task 2 (Category Verification)*: Participants were presented with 3 variations of a category verification task interleaved with an equal number of exposure trials. The three variations of this task included regular, reverse, and

speeded versions. In the regular category verification task, participants were presented with a subordinate-level or basic-level car label, for example "Model Y" (subordinate) or "Other" (basic) for 500 ms, followed by a fixation cross for 250 ms and a picture of a car for 500 ms. If the image and the label matched, the participant pressed a key for SAME. If they did not match, they pressed a key for DIFFERENT. SAME trials included an exemplar from the same class (basic) or same model (subordinate). DIFFERENT trials included an exemplar from a different class (basic) or a different model within the same class (subordinate). In the reverse category verification task, the image was presented before the label (instead of after), and in the speeded version participants completed the regular task but were given only 1 s to respond. On each day participants completed 90 trials of subordinate-level judgments, 90 trials of basic-level judgments, and 90 exposure trials across all three category verification tasks. To equate stimulus exposure across these conditions lures for the DIFFERENT trials were selected from a pool of images of classic cars not previously trained. Exemplars were rotated across days so each exemplar was presented equally throughout the training. Participants were given correct or incorrect feedback for 500 m s following their response; no feedback was give for exposure trials.

3) *Exposure Training*: Within the naming task and verification tasks, either before or after each block (order rotated across blocks and counterbalanced across participants), participants were exposed to the cars in the exposure-only condition. More specifically, either before or after each block of trials participants saw a series of images of cars in the exposure condition. Participants were instructed to pay attention to these images, although no response was required and no feedback was provided. For the exposure trials, a fixation cross was first presented for 250 ms, followed by the presentation of the exposure stimulus for 500 ms. There was an interstimulus interval of 800 ms between each exposure trial. Participants were presented with an equal number of exposure trials as basic and subordinate training trials.

### 4.6. Electrophysiological methods

Scalp voltages were collected with a 128-channel Geodesic Sensor Net™ (Tucker, 1993) connected to an AC-coupled, 128-channel, high input impedance amplifier (200 MΩ, Net Amps™, Electrical Geodesics Inc., Eugene, OR). Amplified analog voltages (0.1–100 Hz bandpass) were digitized at 250 Hz and collected continuously. Individual sensors were adjusted until impedances were less than 40 kΩ. Trials were discarded from analyses if they contained eye movements (vertical EOG channel differences greater than 70 μV) or more than ten bad channels (changing more than 100 μV between samples, or reaching amplitudes over 200 μV). EEG from individual channels that was consistently bad for a given participant was replaced using a spherical interpolation algorithm (Srinivasan et al., 1996).

Stimulus-locked ERPs were baseline-corrected with respect to a 100 ms pre-stimulus recording interval and digitally low-pass filtered at 40 Hz. An average-reference transformation was used to minimize the effects of reference-site activity and accurately estimate the scalp topography of the measured electrical fields. Due to low trial counts, ERPs included all correct and incorrect trials. There was at least an average of 51.9 (SD = 6.7) trials/subject/condition contributing to the average used in the 3 × 2 × 3 × 3 × 2 MANOVA (see below for cell details).

### 4.7. Statistical procedure

#### 4.7.1. Statistical analysis of behavioral measures

To determine whether there was behavioral evidence of an entry-level shift from basic to subordinate processing across training, measures of reaction time for the category verification and matching tasks were entered into a 3 × 6 MANOVA with 3 levels of category level (basic, subordinate, exposure) and 6 levels of training day (day 1, day 2…day 6). In addition, d' analyses were conducted on the pre, post, and 1-week post sequential matching tasks to determine changes in discriminability after training. Planned paired comparisons were used to determine significant d' changes from the pre-test to both post-tests across training and generalization conditions.

#### 4.7.2. Statistical analysis of electrophysiological measures

Electrophysiological analyses of each individual component of interest (N170; N250) was analyzed using separate 3 × 2 × 3 × 3 × 2 MANOVAs including 3 levels of test (pre-test, post-test, 1 week post-test), 2 levels of stimulus presentation order (first, second), 3 levels of categorization training (basic, subordinate, exposure), 3 levels of condition (trained species/trained exemplars, trained species/novel exemplars, and untrained species), and 2 hemispheres (right, left).

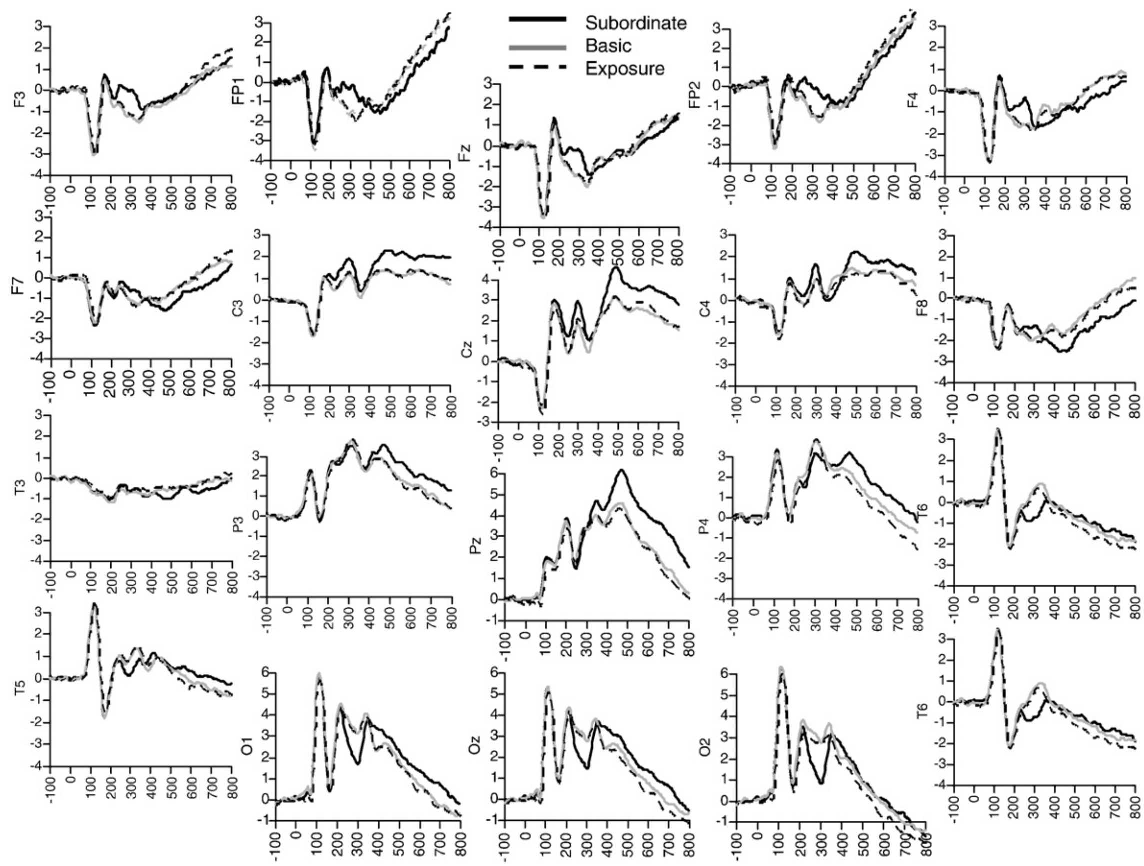
Mean amplitude was calculated within each window of interest for each participant. Statistical analyses were conducted on these means. The channels were selected by identifying the electrode locations in the right and left hemisphere with the largest N170 and N250 across all conditions (channels 70 and 90, between standard locations and O1/O2 and T5/T6). Analyses were conducted on the mean amplitude of averaged ERPs for the N170 (155–211 ms after stimulus onset) and the N250 (230–330 ms after stimulus onset) across these channels and the seven immediately adjacent channels within each hemisphere. Channels of interest were averaged within each hemisphere.

## Acknowledgments

This research was supported by the James S. McDonnell Foundation, the Temporal Dynamics of Learning Center (NSF Grant #SBE-0542013), a grant to T.C., from the National Institute of Mental Health (MH64812) and a grant to J.T. from Natural Sciences and Engineering Research Council of Canada. The authors thank members of the Perceptual Expertise Network for relevant discussion, and to C. DeBuse, B. Woroch, and B. Young for technical and research assistance.

## Appendix A. Sampling of waveforms from the full montage

Pictured is a sampling of waveforms from the full montage (Extended 10–20 system) showing the response to the subordinate-, basic-, and exposure-trained categories during the post-test.





REFERENCES

Anderson, B., Mruczek, R.E.B., Kawasaki, K., Sheinberg, D.L., in press. Effects of familiarity on neural activity in monkey inferior temporal lobe. Cereb. Cortex (Electronic publication ahead of print). doi:10.1093/cercor/bhn015.

Carmel, D., Bentin, S., 2002. Domain specificity versus expertise: factors influencing distinct processing of faces. Cognition 83, 1–29.

Collin, C.A., McMullen, P.A., 2005. Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. Percept. Psychophys. 67, 354–364.

Eimer, M., 2000. The face-specific N170 component reflects late stages in the structural encoding of faces. Neuroreport 11, 2319–2324.

Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology 45, 152–170.

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2006. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cereb. Cortex 16, 1631–1644.

Gauthier, I., Tarr, M.J., 1997. Becoming a "Greeble" expert: exploring mechanisms for face recognition. Vision Res. 37, 1673–1682.

Gauthier, I., Williams, P., Tarr, M.J., Tanaka, J., 1998. Training 'greeble' experts: a framework for studying expert object recognition processes. Vision Res. 38, 2401–2428.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C., 1999. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat. Neurosci. 2, 568–573.

Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. 36 (2), 191–197.

Gauthier, I., Curran, T., Curby, K., Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. Nat. Neurosci. 6, 428–432.

Homa, D., Vosburgh, R., 1976. Category breadth and the abstraction of prototypical information. J. Exp. Psychol: Human Learning and Memory. 2, 322–330.

Jiang, X., Bradley, E., Rini, R.A., Zeffiro, T., VanMeter, J., Riesenhuber, M., 2007. Categorization training results in shape- and category-selective human neural plasticity. Neuron. 53, 891–903.

Luck, S.J., Hillyard, S.A., 1994. Electrophysiological correlates of feature analysis during visual search. Psychophysiology 31, 291–308.

Mack, M.L., Wong, A.C.-N., Gauthier, I., Tanaka, J.W., Palmeri, T.J., 2007. Unraveling the time-course of perceptual categorization: Does fastest mean first? To appear in the Proceedings of the Twenty-Ninth Annual Meeting of the Cognitive Science Society.

Posner, M.I., Keele, S.W., 1968. On the genesis of abstract ideas. J. Exp. Psychol. 77, 353–363.
Riesenhuber, M., Poggio, T., 2000. Models of object recognition. Nat. Neurosci. 3, 1199–1204.
Riesenhuber, M., Poggio, T., 2002. Neural mechanisms of object recognition. Curr. Opin. Neurobiol. 12, 162–168.
Rossion, B., Gauthier, I., Goffaux, V., Tarr, M.J., Crommelinck, M., 2002. Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. Psychol. Sci. 13, 250–257.
Rossion, B., Kung, C.C., Tarr, M.J., 2004. Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. Proc. Natl. Acad. Sci. 101, 14521–14526.
Sagiv, N., Bentin, S., 2001. Structural encoding of human and schematic faces: holistic and part-based processes. J. Cognit. Neurosci. 13, 937–951.
Schweinberger, S.R., Pickering, E.C., Burton, M., Kaufmann, J., 2002. Human brain potential correlates of repetition priming in face and name recognition. Neuropsychologia 40, 2057–2073.
Schweinberger, S.R., Huddy, V., Burton, M., 2004. N250r: a face-selective brain response to stimulus repetitions. Neuroreport 15, 1501–1505.
Scott, L.S., Tanaka, J.W., Sheinberg, D.L., Curran, T., 2006. A reevaluation of the electrophysiological correlates of expert object processing. J. Cognit. Neurosci. 18, 1453–1465.
Simson, R., Vaughan, H.G., Ritter, W., 1977. The scalp topography of potentials in auditory and vidual discrimination tasks. Electroencephal. Clin. Neurophysiol. 42, 529–535.
Srinivasan, R., Nunez, P.L., Silberstein, R.B., Tucker, D.M., Cadusch, P.J., 1996. Spatial sampling and filtering of EEG with spline-Laplacians to estimate cortical potentials. Brain Topogr. 8, 355–366.
Tanaka, J.W., Curran, T., 2001. A neural basis for expert object recognition. Psychol. Sci. 12, 43–47.
Tanaka, J.W., Curran, T., Sheinberg, D.L., 2005. The training and transfer of real-world perceptual expertise. Psychol. Sci. 16, 145–151.
Tanaka, J.W., Curran, T., Porterfield, A.L., Collins, D., 2006. Activation of preexisting and acquired face representations: the N250 event-related potential as an index of face familiarity. J. Cognit. Neurosci. 18, 1488–1497.
Tucker, D.M., 1993. Spatial sampling of head electrical fields: the geodesic sensor net. Electroencephalogr. Clin. Neurophysiol. 87, 154–163.
Williams, R.N., and Tanaka, J.W., in press. Transfer and interference effects in perceptual expertise. In Learning and Generalization, M. Banich .ed. Lawrence Erlbaum Press, NJ.